

DEN TILFÆLDIGE UDVÆLGELSE METODE I STATISTIKKEN.

Af Adolph Jensen.

Ved Anvendelsen af den repræsentative Metode i Statistikken kan man benytte mange forskellige Fremgangsmaader ved Udvælgelsen af den Del af den statistiske Masse, som man vil gøre til Genstand for direkte Undersøgelse. De forskellige Fremgangsmaader vil dog alle kunne indordnes under et af de to Hovedprincipper, der karakteriseres ved Benævnelserne: den „tilfældige“ og den „bevidste“ Udvælgelse af Delmassen. (Undertiden benyttes en Kombination af begge disse Principper).

Ved den tilfældige Udvælgelse sikrer man sig Delmassens repræsentative Karakter ved af Totalmassens Enheder at udtage en Brøkdelen efter et eller andet mekanisk Princip, som er uden Sammenhæng med Undersøgelsens Genstand og Formaal, idet det blot iagttages, at Chancen for at blive udtaget er ens for samtlige Enheder i Totalmassen. Hvis denne Forudsætning er tilstede, kan Sikkerheden naturligvis bestemmes ad matematisk Vej.

I den praktiske Statistik, hvor det gælder om i størst muligt Omfang at spare Tid, Arbejde og Penge, vil man i mange Tilfælde med Fordel kunne anvende den tilfældige Udvælgelse paa den Maade, at man i Stedet for enkeltvis at udtage de Enheder, der skal danne Delmassen, udtager Grupper af Enheder. I Stedet for at undersøge hvert 10de Medlem i Landets Sygekasser, tager man samtlige Medlemmer i hver 10de Sygekasse. Herved faar man imidlertid en mindre Spredning af Delmassens Enheder, og Sandsynligheden for at faa en tilstrækkelig repræsentativ Delmasse er derfor ikke den samme som naar Enhederne udtages enkeltvis. Dette spiller nu ikke nogen synderlig Rolle, hvis der kan udtages et stort Antal Grupper, men i Praksis vil man ofte være nødt til at nøjes med at danne Delmassen af saa faa Grupper, at Virk-

ningerne af Forskelligheder i Gruppernes Sammensætning ikke kan ventes at blive udlignede.

Den mindre fuldkomne Spredning, der fremkommer ved gruppevis Udtagning af Delmassen, kan man imidlertid bøde paa ved at forlade den tilfældige Udvælgelses Princip og gaa over til den bevidste Udvælgelse. I Stedet for at udtage hver 10de Sygekasse efter en mekanisk Regel (f. Eks. efter alfabetisk Rækkefølge), udvælger man en Tiendedel af Kasserne med udtrykkeligt Henblik paa at tilvejebringe en Delmasse, der er saa repræsentativ som muligt. De Grupper, man vælger, maa da være saadanne, som med Hensyn til Kvantiteter eller Kvaliteter, der allerede er kendte, tilsammentagne frembyder samme (eller meget nær samme) Gennemsnit eller Forhold som Totalmassen. Og her vil man naturligvis saa vidt muligt vælge saadanne Kriterier, som har Sammenhæng med Undersøgelsens Genstand.

Metoden kan naturligvis varieres i det uendelige, og der findes en Række Eksempler paa, at den har ført til fortrinlige Resultater. Ved Siden heraf er der dog ikke faa Tilfælde, hvor den bevidste Udvælgelses Metode har skuffet Forventningerne, og disse Skuffelser har undertiden avlet en Pessimisme, som er resulteret i en principiel Afstandtagen fra overhovedet at anvende repræsentative Undersøgelser som Erstatning for fuldstændig statistisk Bearbejdelse af Totalmassen. Jeg har imidlertid taget mig for at undersøge en Del af disse Tilfælde, og jeg har altid kunnet konstatere, at det uheldige Udfald ikke skyldes Metoden som saadan, men den fejlagtige Maade, hvorpaa den er blevet anvendt.¹⁾

Men hvis man nu ser bort fra saadanne Tilfælde, hvor der er begaaet Fejl ved Udvælgelsen af de Grupper, der danner Delmassen, bliver det jo et vigtigt Spørgsmaal, hvilken Grad af Nøjagtighed der kan tillægges Resultaterne af en saadan Undersøgelse, og ved hvilke Midler man kan forøge Nøjagtigheden. Dette Spørgsmaal har Professor Arthur Bowley nylig gjort til Genstand for meget indgaaende Behandling, og det Resultat han er naaet til, vil blive refereret i det følgende, men først skal vi gennem et Eksempel vise, hvorledes Problemet stiller sig i Praksis.

¹⁾ Adolph Jensen: The Representative Method in Practice. (Nordisk Statistisk Tidsskrift, 1925, Side 504 ff.; se særlig Afsnittet „Criticism“ Side 554—562).

Ved den Landbrugstælling, der foretoges i Danmark i 1923, blev der for hver enkelt Ejendom tilvejebragt en Række Oplysninger ud over de sædvanlige. Da en intensiv Bearbejdelse af dette store Materiale vilde udkræve større Penge-midler, end der stod til Raadighed, og da en rent summarisk Optælling vilde være lidet tilfredsstillende, valgte man at foretage Bearbejdelsen i to Afsnit, idet man først foretog en simpel Optælling af samtlige Kolonner paa Skemaerne og dernæst for en vis Brøkdel af Materialet en dybere gaaende Bearbejdelse, hvorved de enkelte Data kombineredes paa forskellig Maade.¹⁾

De sidste Afsnit af Bearbejdelsen foretoges efter den repræsentative Metode med bevidst Udvalgelse af Delmassen, hvis Omfang paa Forhaand fastsattes til ca. en Femtedel af samtlige Landbrugsejendomme. Ved Udvælgelsen af Delmassen benyttede man Kommunerne som Grupper, og da Landet har omtrent 1300 Landkommuner, kom Delmassen til at bestaa af 260 Grupper. Da man imidlertid ønskede, at Undersøgelsen ogsaa territorialt set skulde være repræsentativ, blev Kommunerne udvalgte amtsvis, saaledes at der i Virkeligheden dannedes 22 Delmasser. Antallet af udvalgte Kommuner varierede naturligvis fra Amt til Amt, men det siger sig selv, at Antallet i intet Tilfælde var betydeligt, og der maatte derfor anvendes den størst mulige Omhu paa Udvælgelsen, saaledes at hver enkelt af de 22 Delmasser — trods det ringe Antal Grupper — kom til at frembyde meget nær samme Gennemsnitsegenskaber som Amtets samtlige Kommuner under eet.

Som det principale Kriterium ved Udvælgelsen valgte man Forholdet mellem Antallet af Malkekøer og Størrelsen af det landbrugsmæssigt benyttede Areal. Indenfor hvert Amt blev der for hver Kommune foretaget en Beregning af Antallet af Køer pr. 100 Tdr. Land Landbrugsareal. Derefter ordnedes Kommunerne i Rækkefølge efter de saaledes udregnede Forholdstal, og af disse Rækker udtoges i første Omgang hver femte Kommune. For at kontrollere, at de forskellige Egne af Amtet fik en nogenlunde ligelig Repræsentation, blev de udtagne Kommuner indtegnede paa et Kort, og hvor det skønedes fornødent, foretoges Ombytning af enkelte Kommuner med andre af saa vidt muligt samme „Ko-Tæthed“.

For at give en Forestilling om, hvor nødvendigt det er at anvende megen Omhu paa Gruppernes Udvalgelse efter det

¹⁾ Statistiske Meddelelser, 4. Række, 78. Bind, 2. Hæfte.

valgte Kriterium, skal anføres, at i et Amt som Præstø, hvor dog Jordboniteten og andre Naturforhold er ret ensartede, varierer Kommunernes Ko-Tæthed saa betydeligt som mellem 8,6 og 35,5 med 28,2 som Amts-Gennemsnit. Spredningens Karakter fremgaar af følgende:

Under 80 % af Gennemsnittet	7	Kommuner
80—90 %	7	—
90—100 %	20	—
100—110 %	22	—
110—120 %	15	—
120 % af Gennemsnittet eller derover . .	5	—

Hvor Fordelingen har en saa symmetrisk Form som her, vil Udvælgelse efter den foran beskrevne Fremgangsmaade i Almindelighed give et i Relation til det benyttede Kriterium tilfredsstillende Resultat, selv om Antallet af udtagne Grupper er lille, absolut set. Men selv om en Delmasse giver et ganske nøjagtigt Billede af Totalmassen med Hensyn til en enkelt Egenskab, har man jo ingen Sikkerhed for, at den er tilstrækkeligt repræsentativt med Hensyn til andre Egenskaber. For at vende tilbage til Eksemplet med Præstø Amt, frembyder Rodfrugtdyrkingen her en Ejendommelighed, som kræver særlig Opmærksomhed. For Amtet som Helhed er 13,1 % af Landbrugsarealet anvendt til Rodfrugter; Kommune-Gennemsnittene varierer mellem 4,3 og 19,5, men Fordelingen er usymmetrisk, idet der er to Maksima, det ene omkring 11, det andet omkring 17. Grunden hertil er, at der i visse Egne af Amtet dyrkes Sukkerroer. Hvis man ordner Kommunerne efter aftagende Rodfrugtprocent, faar man derfor en Række, som i væsentlig Grad afviger fra Rækken efter aftagende Ko-Tæthed, og den Delmasse, der alene dannes under Hensyntagen til det sidstnævnte Kriterium, vil derfor være mindre tilfredsstillende set i Forhold til det førstnævnte. Endnu mere grelt stiller Forholdet sig med Hensyn til Engarealet, hvis Fordeling jo i det væsentlige er bestemt ved Naturforhold. For Præstø Amts Kommuner viser Eng-Procenten (Engareal i Procent af samlet Landbrugsareal) nedenstaaende Fordeling, som især er ejendommelig, naar den sammenholdes med den ovenfor anførte Fordeling efter Ko-Tæthed.

Under 80 % af Gennemsnittet	38	Kommuner
80—90 %	5	—
90—100 %	7	—
100—110 %	2	—
110—120 %	3	—
120 % af Gennemsnittet eller derover . .	21	—

Det er klart, at Iagttagelser som de her nævnte indebærer en Opfordring til at prøve Materialet i forskellige Retninger, før man tager endelig Bestemmelse om de Delmasser, man vil bearbejde. Som Vejledning ved dette Forarbejde anbefaler det sig at undersøge, hvilken Korrelation der bestaar mellem de forskellige Forhold, som der er Mulighed eller Rimelighed for at benytte som Kriterier ved Udvælgelsen af Delmassen.

Ved Siden af det principale Kriterium, der anvendtes ved Landbrugsstatistikken 1923, nemlig Ko-Tætheden, laa det nær at have Opmærksomheden henvendt paa følgende Momenter: den samlede Hornkvægbestand, Hestebestanden, Svinebestanden, Kornarealet, Rodfrugtarealet, Græsarealet i Agermarken og Engarealet — alt i Forhold til Kommunens samlede Landbrugsareal. Den indbyrdes Afhængighed mellem disse forskellige Rækker af Forholdstal kan man danne sig et Billede af ved at beregne Korrelationskoefficienten

$$r = \frac{\Sigma xy}{n \sqrt{\frac{\Sigma x^2}{n}} \sqrt{\frac{\Sigma y^2}{n}}}$$

hvor x og y betegner Afvigelserne fra Gennemsnittene og n betegner Antallet. (De to Rodstørrelser i Nævneren er Middelfejlene paa x og y). For Korrelationskoefficienten (r) mellem Ko-Tætheden og de øvrige ovenfor nævnte Momenter finder man for Præstø Amt følgende Værdier:

Køer og Hornkvæg	0,924
— - Heste	0,703
— - Svin	0,689
— - Kornareal	0,697
— - Rodfrugtareal	0,438
— - Græsareal i Agermarken	0,577
— - Engareal	÷ 0,531

Som det var at vente, er der næsten fuldstændig Korrelation mellem Ko-Bestanden og den samlede Bestand af Hornkvæg. Køer og Engareal varierer i modsat Retning (Korrelationskoefficienten negativ). Mindst god er Korrelationen mellem Køer og Rodfrugtareal, hvor r har den numerisk mindste Værdi.

En Prøve som denne kan naturligvis ikke give andet end et Fingerpeg. Det maa altid bero paa et Skøn, hvor daarlig Korrelationen mellem to Rækker bør være for at begrunde, at man medtager det paagældende Forhold som Kriterium ved

Udvælgelsen af Delmassen. Ved Landbrugsstatistikken 1923 medtog man alle de nævnte Momenter, uagtet det sikkert vilde have været fuldt forsvarligt at bortse i hvert Fald fra den samlede Hornkvægbestand.

Udvælgelsen af Delmassen efter en Flerhed af Kriterier kan naturligvis ske paa mange Maader. Hvis Totalmassen bestaar af et i Forhold til Antallet af Kriterier stort Antal Grupper, kan man gaa ganske mekanisk til Værks. Hvis man f. Eks. havde en Totalmasse bestaaende af 1250 Grupper, hvoraf man med Benyttelse af 3 Kriterier vilde udtage en Delmasse paa 20 %, kunde man gaa frem paa følgende Maade. De 1250 Grupper ordnes efter 1ste Kriterium og deles derefter i 10 lige store Afdelinger. De 250 Grupper indenfor hver af de 10 Afdelinger ordnes efter 2det Kriterium, hvorefter de deles i 5 lige store Underafdelinger. De 25 Grupper indenfor hver Underafdeling ordnes efter 3die Kriterium, og af det saaledes ordnede Materiale udtages sluttelig hver 5te Gruppe, ialt 250 Grupper, som da vil udgøre en Delmasse, som har en ikke ringe Sandsynlighed for med Hensyn til alle tre Kriterier at være repræsentativ i Forhold til Totalmassen.

Den her beskrevne Fremgangsmaade — man kunde kalde den Lagdelingsmetoden — lader sig imidlertid ikke anvende, hvor Antallet af Grupper er lille i Forhold til Antallet af Kriterier. Ved Landbrugsstatistikken 1923 foretoges den Ombytning af Grupper, der var nødvendig for at tilfredsstille de sekundære Kriterier, saa at sige „paa Øjemaal“. For Øerne og Jylland og for hele Landet under eet fremgaar Resultatet af nedenstaaende Tal, der angiver Delmassens Procentandel af Totalmassen.

	Øerne	Jylland	Hele Landet
Antal Landbrugsejendomme	20,6	20,9	20,8
Samlet Landbrugsareal	20,8	20,7	20,8
Kornareal	21,0	20,7	20,8
Rodfrugtareal	21,0	20,8	20,9
Grøntfoder- og Græsningareal i Agermarken	21,1	20,8	20,9
Engareal	19,5	20,9	20,7
Antal Heste	21,0	20,6	20,8
Antal Køer	20,8	20,8	20,8
Antal Stkr. Hornkvæg i alt	20,8	20,9	20,9
Antal Svin	20,8	21,1	21,0

Delmassen omfatter for hele Landet under eet 40 700 af Totalmassens 195 700 Landbrugsejendomme, altsaa paa det nærmeste 21 %, og det ses af ovenstaaende Tal, at i meget

nær samme Forhold var de forskellige Arealanvendelser og Kreaturarter repræsenteret. Ogsaa for de enkelte Amter samlede Forholdstallene sig nær omkring 21 0/0, om der end her og der, navnlig for Engarealets Vedkommende, var en Del Afvigelser fra Gennemsnittet, som i nogen Grad kan paavirke Sikkerheden af Undersøgelsens Resultater, for saa vidt angaar de enkelte Egne af Landet.

Medens den Nøjagtighed, hvormed Resultaterne beregnes, naar man ved Dannelsen af Delmassen benytter den tilfældige Udvalgelse, er forholdsvis let at maale, gælder det samme ikke, naar Delmassen dannes ved bevidst Udvalgelse. Ja, indtil for ganske nylig har det endog været den almindelige Opfattelse, at man her maatte nøjes med den Tryghedsfølelse, som det giver, at der er anvendt al mulig Omhu paa at tilvejebringe en Delmasse, der ligner Totalmassen paa de Punkter, der har Betydning for den paagældende Undersøgelse; derimod har man ment, at en eksakt Maaling af Nøjagtigheden var udelukket. Denne Opfattelse deltes ogsaa af Professor Bowley, indtil han for et Par Aar siden som Medlem af en af det internationale statistiske Institut nedsat Kommission kom til at beskæftige sig nærmere med nogle Specialspørgsmaal, hvorved han førtes ind paa den Tanke, at Korrelations-Teorien her — som paa saa mange andre Punkter — kunde yde den praktiske Statistik værdifuld Støtte. Resultaterne af Bowleys Arbejde er nedlagt i en Afhandling, der er trykt som Bilag til den før nævnte Kommissions-Betænkning.¹⁾ Uden at komme ind paa Detaillerne i Bowleys Udredning, skal jeg her gengive Hovedresultatet, og ved et Par Eksempler forsøge at paavise den praktiske Værdi af hans Formel for Maaling af Nøjagtigheden.

Det Problem, der foreligger, kan i sin Almindelighed karakteriseres saaledes.

Af en Totalmasse paa N Grupper, der tilsammen indeholder A Enheder, udtages en Delmasse paa n Grupper med tilsammen a Enheder, saaledes at Summen af de n Grupper med Hensyn til et eller flere Kriterier („Kontroller“) frembyder samme Forhold som Summen af de N Grupper. Hvilken Usikkerhed klæber der da ved at slutte, at et eller andet Gen-

¹⁾ A. L. Bowley: Measurement of the Precision attained in Sampling. (Bulletin de l'Institut International de Statistique. Tome XXII. 1. livraison. Roma 1926).

nemnsnit eller Forhold, om hvilket man ved, at det for Delmassens a Enheder har Værdien x , for Totalmassens A Enheder har Værdien $X = Ax:a$?

Vi vil først betragte det Tilfælde, at der ved Udvælgelsen af Delmassen kun anvendes en enkelt Kontrol, og derefter vise hvorledes Spillerummet for Usikkerheden paavirkes derved, at der samtidig anvendes flere Kontroller.

For det første er det indlysende, at Usikkerheden maa være des mindre, jo flere Grupper Delmassen bestaar af. Usikkerheden aftager imidlertid ikke i direkte Forhold til n , men — ligesom ved den tilfældige Udvælgelse — med Kvadratrodten af n .

Bowley har nu paavist, at foruden Antallet af Grupper i Delmassen er der tre andre Momenter, der paavirker Spillerummet for Usikkerheden, saaledes at Middelfejlen udtrykkes ved et Produkt af fire Faktorer. Den første af disse er, som

ovenfor nævnt, $\frac{1}{\sqrt{n}}$; de tre andre knytter sig henholdsvis til det søgte Gennemsnit (x), til Fordelingen af Delmassens Enheder (a) og til den benyttede Kontrol. Vi skal betragte de tre sidstnævnte Momenter hver for sig.

Til hver af Delmassens n Grupper svarer der en Værdi af x , som for den s 'te Gruppe kan betegnes ved x_s . Jo mere de forskellige Værdier af x_s afviger fra Gennemsnitsværdien \bar{x} , des større bliver Spillerummet for Usikkerheden, og Maalet herfor er Middelfejlen paa x , nemlig:

$$\sigma_x = \sqrt{\frac{\sum(x_s - \bar{x})^2}{n}}$$

Men ogsaa Fordelingen af Delmassens Enheder (a) paavirker Usikkerheden, om end i Almindelighed i mindre Grad end x . Bowley viser, at denne Paavirkning maales ved Værdien af

$$\sqrt{1 + \frac{\sigma_a^2}{a^2}}$$

hvor σ_a er Middelfejlen paa a og \bar{a} er Gennemsnittet $\frac{\sum a_s}{n}$.

Endelig paavirkes Spillerummet for Usikkerheden af Korrelationen mellem x og den benyttede Kontrol. Hvis Størrelsen af Kontrollen for Totalmassen er U , kan dens tilsvarende Værdi for den s 'te Gruppe betegnes ved u_s og

Af Tavlen fremgaar det, hvor vigtigt det er at vælge en Kontrol, som har god Korrelation med x , men det ses tillige, at hvis dette ikke er muligt, vil den samme Sikkerhed kunne opnaas ved at forøge Antallet af Grupper i Delmassen. Med Korrelationskoefficient 0,2 og 5 n Grupper opnaar man samme Sikkerhed som med Korrelationskoefficient 0,9 og n Grupper. Korrelationskoefficient 0,5 og 2 n Grupper giver samme Sikkerhed som Korrelationskoefficient 0,8 og n Grupper. Valget af Antallet af Grupper er imidlertid ofte et vigtigt økonomisk Spørgsmaal, idet Omkostningerne ved Delmassens Bearbejdelse kan regnes at være omtrent proportionale med dens Størrelse, og man vil derfor i Praksis vælge Antallet af Grupper ikke alene under Hensyn til Kontrollens bedre eller mindre gode Korrelation med x , men ogsaa tage i Betragtning den Grad af Sikkerhed, som Formaalet med den paagældende Undersøgelse kræver.

Der er imidlertid en tredje Vej, ad hvilken Spillerummet for Usikkerheden kan formindskes. Det er ved at benytte en Flerhed af Kontroller. I Stedet for at udtage Delmassen alene efter Kontrollen U , søger man altsaa en Delmasse, som baade med Hensyn til $U, V, W \dots$ o. s. v. irembyder samme Gennemsnit eller Forhold som Totalmassen. De tilsvarende Værdier for den s 'te Gruppe bliver da $u_s, v_s, w_s \dots$ og Korrelationskoefficienterne med x bliver $r_u, r_v, r_w \dots$. Maalet for Usikkerhedens Spillerum vil imidlertid i saa Fald ikke alene paavirkes af Korrelationen mellem x og Kontrollerne, men ogsaa af Korrelationen mellem disse sidste indbyrdes, $q_{uv}, q_{uw}, q_{vw} \dots$.

Hvis man nu giver det sidste Led i Formlen for μ Formen $\sqrt{R:R_t}$, hvor t er Antallet af Kontroller, vil R og R_t bestemmes ved Determinanterne

$$R = \begin{vmatrix} 1 & r_u & r_v & r_w & \dots \\ r_u & 1 & q_{uv} & q_{uw} & \dots \\ r_v & q_{uv} & 1 & q_{vw} & \dots \\ r_w & q_{uw} & q_{vw} & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{vmatrix} \quad R_t = \begin{vmatrix} 1 & q_{uv} & q_{uw} & \dots \\ q_{uv} & 1 & q_{vw} & \dots \\ q_{uw} & q_{vw} & 1 & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{vmatrix}$$

Hvorledes vil nu Forøgelsen af Kontrollernes Antal paavirke Størrelsen af μ ? Vi saa foran, at hvor der kun er een Kontrol, er $R:R_t = 1 - r_u^2$.

Hvis der er to Kontroller (altsaa $t = 2$), faas af foranstaaende Determinanter

$$R:R_t = \frac{1 + 2 r_u r_v \varrho_{uv} - (r_u^2 + r_v^2 + \varrho_{uv}^2)}{1 - \varrho_{uv}^2},$$

og for højere Værdier af t faar man langt mere indviklede Udtryk for $R:R_t$. Disse Udtryk er imidlertid saa lidet overskuelige, at det vil anbefale sig at simplificere Spørgsmaalet paa den Maade, at man forudsætter samme Korrelation mellem x og samtlige Kontroller og ligeledes samme Korrelation mellem Kontrollerne indbyrdes, altsaa:

$$r_u = r_v = r_w = \dots = r \text{ og } \varrho_u = \varrho_v = \varrho_w = \dots = \varrho.$$

I saa Fald faar man nemlig for $R:R_t$ følgende Udtryk, der er let at overskue og bekvemt at regne med:

$$R:R_t = 1 - \frac{tr^2}{(t-1)\varrho + 1}$$

Ved de Opgaver, der forekommer i Praksis, vil det let kunne tænkes, at Korrelationen mellem x og Kontrollerne ligger omkring $\frac{2}{3}$, og at Korrelationen mellem Kontrollerne indbyrdes grupperer sig nær omkring $\frac{1}{2}$. Hvis man gaar ud fra disse Værdier for r og ϱ , vil Størrelsen af $\sqrt{R:R_t}$ aftage som nedenfor angivet, efterhaanden som Antallet af Kontroller forøges:

$t =$	$\sqrt{R:R_t} =$	$t =$	$\sqrt{R:R_t} =$
1	0,745	6	0,486
2	0,638	7	0,472
3	0,578	8	0,458
4	0,538	9	0,448
5	0,509	10	0,438

Det vil ses, at naar Antallet af Kontroller forøges, aftager $\sqrt{R:R_t}$ i Begyndelsen ret stærkt, men efterhaanden mindre og mindre. Hvis man tænkte sig noget saa urimeligt som at benytte 1000 Kontroller, vilde Værdien af $\sqrt{R:R_t}$ dog ikke gaa længere ned end til 0,334, og for $t = \infty$ bliver $\sqrt{R:R_t} = 0,333$.

Dette vil altsaa sige, at man overfor den Sikkerhedsgrad, hvormed et givet Gennemsnit eller Forhold beregnes, ikke kan vente sig alt for stor Hjælp af at forøge Kontrollernes Antal ud over nogle faa. Ligesom foran ved Vurderingen af

den Indflydelse det øver paa Sikkerheden, om Korrelationen mellem x og Kontrollen er mere eller mindre god, ledes vi her til at tage i Betragtning den foran paaviste Virkning af en Forøgelse eller Formindskelse af det Antal Grupper (n), der indgaar i Delmassen.

Nedenstaaende Tavle angiver en Række af de Faktorer, hvormed μ reduceres ved Kombination af n , $1,5 n$, $2 n \dots \dots 10 n$ og $t = 1, 2, 3 \dots \dots 10$, idet det stadig forudsættes, at r og ρ har de konstante Værdier $\frac{2}{3}$ og $\frac{1}{2}$.

$t =$	n multipliceret med										
	1,0	1,5	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0	10,0
1	0,75	0,61	0,52	0,43	0,37	0,33	0,30	0,28	0,25	0,25	0,24
2	0,64	0,52	0,45	0,37	0,32	0,29	0,26	0,24	0,23	0,21	0,20
3	0,58	0,47	0,41	0,33	0,29	0,26	0,24	0,22	0,20	0,19	0,18
4	0,54	0,44	0,38	0,31	0,27	0,24	0,22	0,20	0,19	0,18	0,17
5	0,51	0,42	0,36	0,29	0,26	0,23	0,21	0,19	0,18	0,17	0,16
6	0,49	0,40	0,34	0,28	0,24	0,22	0,20	0,18	0,17	0,16	0,15
7	0,47	0,39	0,33	0,27	0,24	0,21	0,19	0,18	0,17	0,16	0,15
8	0,46	0,38	0,32	0,27	0,23	0,20	0,19	0,17	0,16	0,15	0,15
9	0,45	0,37	0,32	0,26	0,22	0,20	0,18	0,17	0,16	0,15	0,14
10	0,44	0,36	0,31	0,25	0,22	0,20	0,18	0,16	0,16	0,15	0,14

Det vil eksempelvis ses, at man opnaar samme Formindskelse af Spillerummet for Usikkerheden ved at fordoble Antallet af Grupper som ved at forøge Antallet af Kontroller fra 1 til 5, og at en Forøgelse af n til det tredobbelte giver samme Formindskelse af μ som en Forøgelse af t fra 1 til 10.

De foran gengivne Undersøgelser af den Virkning, som Antallet af Grupper i Delmassen, Korrelationen mellem x og Kontrollerne og disse sidstes Antal øver paa Spillerummet for Usikkerheden, har givet det Hovedindtryk, at det førstnævnte Moment spiller en langt større Rolle end de to andre. De praktiske Konklusioner, man kan drage heraf, skal vi komme tilbage til senere. Men først vil vi gennem et Eksempel vise Anvendelsen af Bowleys Formel paa et konkret Materiale. Vi vælger hertil den foran nævnte landbrugsstatistiske Undersøgelse for Præstø Amt, idet vi tænker os, at vi paa Grundlag af de 14 Kommuner, der danner Delmassen, vil bestemme Antallet af Køer under 6 Aar i Amtets 76 Kommuner.

Landbrugsarealet i de 76 Kommuner udgjorde 251 964 Tdr. Land, i de 14 Kommuner 52 524 Tdr. Land. I de 14

Kommuner fandtes 7632 Køer under 6 Aar, og vi slutter da, at Antallet af Køer under 6 Aar i de 76 Kommuner var $7632 \cdot \frac{251\,964}{52\,524} = 36\,634$. Vi søger nu at bestemme Middelfejlen paa dette Resultat.

Det vil erindres, at Delmassen blev udtaget med Anvendelse af ikke mindre end 8 Kontroller (Køer, Hornkvæg, Heste, Svin, Kornareal, Rodfrugtareal, Græsareal og Engareal). For at undgaa det vidtløftige Arbejde, det er at regne med Determinanter af 7de Orden, vil vi imidlertid gøre den Fiktion, at der kun er anvendt 3 Kontroller, nemlig de tre førstnævnte.

Til Bestemmelse af Formlens 1ste Led har man:

$$n = 14; \frac{1}{\sqrt{n}} = 0,267.$$

For at bestemme 2det Led, Middelfejlen paa x , beregnes først Antallet af Køer under 6 Aar pr. 1000 Tdr. Land i hver af de 14 Kommuner og derefter Gennemsnittet af disse 14 Tal og Summen af de 14 tilsvarende Kvadratafvigelse. Idet Summen af Afvigelseernes Kvadrater er $\sum (x_s - \bar{x})^2 = 6603$, bliver Middelfejlen

$$\sigma_x = \sqrt{\frac{6603}{14}} = 21,73.$$

Til Bestemmelse af 3dje Led har man den Række, der angiver Landbrugsarealets Størrelse i Tdr. Land indenfor hver af de 14 Kommuner. Den gennemsnitlige Størrelse er $\bar{a} = 5752$, Middelfejlen $\sigma_a = 965,5$, hvoraf man finder

$$\sqrt{1 + \frac{\sigma_a^2}{a^2}} = \sqrt{1,06622} = 1,0326.$$

For endelig at kunne bestemme fjerde Led søger vi paa Grundlag af Delmassens 14 Kommuner Korrelationskoefficienterne mellem Køer under 6 Aar og henholdsvis Køer (i alle Aldre), Hornkvæg (af alle Arter) og Heste, medens Korrelationen mellem de nævnte Kontroller indbyrdes forudsættes at være kendt (jfr. foran). De paagældende Korrelationskoefficienter er følgende:

Mellem Køer under 6 Aar og Køer.	r_u	$= 0,948$
— — — — — Hornkvæg	r_v	$= 0,970$
— — — — — Heste	r_w	$= 0,924$

Mellem Køer og Hornkvæg	$q_{uv} = 0,924$
— — - Heste	$q_{uw} = 0,703$
— Hornkvæg og Heste	$q_{vw} = 0,825$

Med Benyttelse af disse Værdier beregnes $R = 0,003$ og $R_t = 0,129$, hvoraf Værdien af Formlens fjerde Led

$$\sqrt{R:R_t} = 0,152.$$

Produktet af de nu beregnede 4 Led giver Middelfejlen paa Antallet af Køer under 6 Aar pr. 1000 Tdr. Land, og da de 76 Kommuner Landbrugsareal er 251 964 Tdr. Land, faar vi

$$\mu = 0,267 \times 21,73 \times 1,0326 \times 0,152 \times 251\,964 = 231.$$

Vi har altsaa beregnet Antallet af Køer under 6 Aar i de 76 Kommuner til 36 634 med en Nøjagtighed, der er bestemt derved, at Middelfejlen er 231.

Hertil er nu imidlertid at bemærke, at dette overordentlig fine Resultat skyldes to særlige Omstændigheder. For det første vil det erindres, at Delmassen er udtaget med Anvendelse af 8 Kontroller, medens vi har fingeret, at der kun er anvendt 3. Dette har dog sandsynligvis kun lidt at betyde lige overfor den foreliggende Opgave, idet det ikke kan antages, at Medtagelsen af de 5 sidste Kontroller (Svin, Kornareal, Rodfrugtareal, Græsareal og Engareal) i synderlig Grad har forøget den repræsentative Karakter af Delmassen i Henseende til Antallet af Køer over 6 Aar, som allerede er opnaaet ved Anvendelse af de tre Kontroller Køer, Hornkvæg og Heste. Derimod er det naturligvis afgørende for den lille Værdi af μ , at vi har været i det usædvanlig heldige Tilfælde at kunne anvende Kontroller, hvis Korrelationskoefficienter overfor x alle tre ligger over 0,9. Havde vi valgt en anden Opgave, hvor Korrelationen mellem Kontrollerne og x havde været mindre, vilde vi have faaet en større Middelfejl.

Hovedsagen er imidlertid, at medens man hidtil har maattet nøjes med at kunne betegne Resultatet af en Undersøgelse med Benyttelse af bevidst Udvælgelse af Delmassen som „tilnærmelsesvis rigtigt“, har Bowleys Formel givet os Midler i Hænde til at fastslaa de sandsynlige Grænser for den Fejl, vi risikerer at begaa ved at regne med dette tilnærmede Resultat.

Hovedresultaterne kan sammenfattes saaledes. Spillerummet for Usikkerheden vil være des mindre, jo flere Grupper der

indgaar i Delmassen, jo bedre Korrelation der er mellem Kontrollerne og det søgte, og jo flere Kontroller der anvendes. I Almindelighed vil der dog opnaas mere i Henseende til Sikkerhed ved at forøge Antallet af Grupper end ved at forøge Antallet af Kontroller, især naar disse sidste er godt korrelerede indbyrdes. Endelig vil Forøgelsen af Sikkerheden blive mindre og mindre for hver ny Kontrol der tilføjes.

I Forbindelse med disse almindelige Regler bør imidlertid følgende have i Erindring.

Da en Forøgelse af Antallet af Grupper i Almindelighed vil være ensbetydende med en proportional Forøgelse af Delmassens Enheder, hvorved Bearbejdelsesomkostningerne stiger tilsvarende, bør Valget af Grupper og Kontroller ske under omhyggelig Afvejen af det dobbelte Hensyn til Sikkerhed og Økonomi. Herved bør det tages i Betragtning, at medens det kan være berettiget at bringe store Ofre af Tid, Arbejde og Penge for at opnaa den størst mulige Sikkerhed, naar Sikkerhedsgraden ikke lader sig maale, stiller Sagen sig anderledes, naar man, saaledes som her, er i Stand til at angive Grænserne for Usikkerheden. Den Nøjagtighed, der kræves i de enkelte Tilfælde, afhænger jo af Undersøgelsens Art og Formaal, og det er urimeligt at anvende Omkostninger paa at tilvejebringe en Nøjagtighed, som i det givne Tilfælde er ufornöden.

Spørgsmaalet om Valget af Antallet af Kontroller har imidlertid en Side, som endnu ikke er berørt.

I hele den foregaaende Udvikling er det forudsat, at man staar overfor en saa usammensat Opgave som at finde Værdien af en enkelt Størrelse, et Gennemsnit eller et Forhold i Totalmassen. Herpaa — og ikke paa andet — sigter Bowleys Formel; den angiver Størrelsen af Middelfejlen paa et enkelt Tal, som man søger at bestemme. Naar det foran er paavist, at der opnaas forholdsvis lidt ved at forøge Antallet af Kontroller ud over en ret snæver Grænse, maa dette da ogsaa forstaaes med den nævnte Forudsætning in mente. Men i den praktiske Statistik vil Opgaven som oftest være mere sammensat. Landbrugsstatistikken af 1923 blev jo ikke iværksat alene for at bestemme Antallet af Køer under 6 Aar, men for at søge en Mangfoldighed af andre Forhold belyst. Der var med andre Ord ikke et enkelt X , men en lang Række af meget forskelligartede Ubekendte. Men heraf følger, at selv om det til Bestemmelse af Antallet af unge Køer vilde være tilstrækkeligt at have 3, 2 eller maaske endog kun 1 Kontrol,

maatte Undersøgelsens videre Formaal kræve, at Delmassen blev udtaget under Hensyntagen til en Række andre Momenter end dem, der har Tilknytning til Kobestanden. Det er klart, at jo flere Kontroller man medtager, des mere nærmer man sig den ideale Delmasse, der giver et i alle Henseender nøjagtigt Billede af Totalmassen. En Art Maalestok for, hvor nær man ved Landbrugstællingen 1923 kom til dette Ideal, haves i de foran anførte Tal, der viser, at med Hensyn til alle de 8 benyttede Kriterier rummede Delmassen meget nær 21 % af Totalmassen.
